DETAILED ACTION

This action has been made non-final merely due to 35 USC 101.

Claim Rejections - 35 USC § 101

1. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 1-5 and 17-20 are rejected under 35 U.S.C. 101 because:

The claimed invention is directed to non-statutory subject matter.

As per the claims, the language "computer readable medium" do not transform the claimed subject matter into statutory subject matter. The present invention does not disclose anything. There is absolutely no mention or support thereof of for a medium any kind within the specification. Examiner must therefore construe a computer readable medium as a non-statutory signal such as a carrier wave.

NOTE:

Claims that recite nothing but the physical characteristics of a form of energy, such as a frequency, voltage, or the strength of a magnetic field, define energy or magnetism, per se, and as such are nonstatutory natural phenomena. O'Reilly, 56 U.S. (15 How.) at 112-14.

Response to Arguments

Applicant's arguments filed 07/20/2011 have been fully considered but they are not persuasive.

Argument (page 10 section I. and II. & III. referring back to I.):

- "Though the sections of Spyros cited in the Office Action briefly describe classifying audio segments by gender, estimating gender-dependent band-specific models from training, detecting channel changes, detecting gender changes, using a gender-independent model to detect channel changes, and training a gender-independent, context-independent model using labeled training data, none of the cited sections or any other section of Spyros discloses "creating a gender-independent phoneme model when the difference between the compared female phoneme model and the corresponding male phoneme model is less than a predetermined value" as set forth in Assignee's claim 1. There is nothing in Spyros that describes taking any action "when the difference between the compared female phoneme model and the corresponding male phoneme model is less than a predetermined value." Kanevsky also fails to disclose this limitation. Kanevsky is directed to off- line detection of textual topical changes, and does not even describe "creating a gender-independent phoneme model.

Since none of Neti, Spyros, and Kanevsky disclose "creating a gender-independent phoneme model when the difference between the compared female phoneme model and the corresponding male phoneme model is less than a predetermined value" the rejection of claim 1 should be withdrawn"

Response to argument:

Examiner disagrees and maintains the use of Neti and Spyros. Initially, Neti teaches a method of constructing a gender-dependent speech recognition model includes the steps of providing training data of a known gender, aligning the training data, tagging the training data with a gender to create gender-tagged data, determining a gender question at a node to determine gender dependence of the gender-tagged data, determining a phonetic context question at the node to determine phonetic context dependence of the gender-tagged data, determining a highest value of an evaluation function between the gender dependence and the phonetic context dependence to determine which dependence is a dominant dependence, splitting the data of the dominant dependence into child nodes according to likelihood criteria, comparing the highest value with a threshold value to determine if additional splitting is necessary, repeating theses steps for each child node until the highest value is below the threshold value and counting the nodes having gender dependence to determine an overall gender dependence level. A gender-dependent speech recognition system includes an input device for inputting speech to a preprocessor. The preprocessor converts the speech into acoustic data, and a processor for identifies gender-dependent phone state

models and phone state modes common to both genders. The phone state models are stored in a memory device wherein the processor recognizes the speech in accordance with the phone state models (Neti Abstract & Fig. 2 and 4 below).

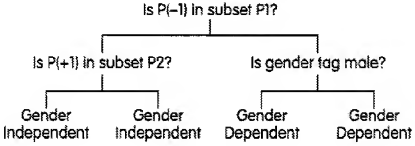

Fig. 2

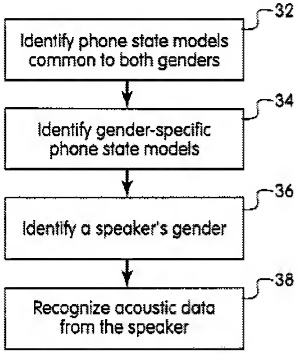

Fig. 4

After further analysis, it is clear that Neti teaches that there is a comparison taking place to determine a commonality or difference between speech to determine whether it is gender dependent or gender independent (i.e. male or female). While, this analysis is clearly demonstrated in Neti having a gender *dependent* model created, Examiner agrees that it is *not* clear whether an independent model is created. In other words it seems as if in Neti, once gender independence is found, further processing stops, which can prevent critical speech from being processed.

Examiner would also like to point out that, in conventional gender dependent systems a complete male model and a complete female model is required (Neti Col 5 lines 32-42).

For instance, the step of identifying gender may include comparing Gaussian prototypes to a codebook of Gaussian prototypes to determine gender. The step of identifying gender specific phone state models further comprises the step of asking a gender question at a node to determine gender dependence of the acoustic data (Col. 1 line 66 – Col. 2 line 9 & Fig. 2).

More specifically, Neti teaches well known uses in gender identification, such as Acoustic training data is typically divided into 10 msec segments called frames. Each frame is represented by an acoustic feature vector. For example, a 1 sec duration would contain 100 frames. The sequence of acoustic vectors are aligned to the phonetic transcription of the utterance. Each phone has three states. After alignment, each state has a subset of acoustic vectors associated with each state which can be modeled by Gaussian prototypes. Distances are computed from each vector in an utterance to the corresponding gender-specific sub-phone models as follows: Let x(t) denote the acoustic vector at time t. .mu..sup.j.sub.L(t) (i) is the mean of the jth prototype of the context-dependent state L(t) corresponding to the

alignment of x(t) and models of gender i. Then, the distance of x(t) to gender i=1, 2 is defined as:

$$dist(i) = \sum_{t} \min_j 1/N \left\| x(t) - \mu^{j}_{L(t)}(i) \right\|^2$$

where .parallel.x(t)-.mu..sup.j.sub.L(t) (i).parallel. This represents the Euclidian distance between two vectors and N is the dimension of the vector. The gender i corresponding to the lowest value of dist(i) is used as the gender tag for the utterance. This method requires sufficient coverage of the phone set in the test utterance and yields poorer gender ID for short utterances. As an example, SDGid is about 62% accurate for 3-8 second utterances (Col. 6 lines 24-50).

This means that an improvement of Neti is required to account for only 62% accuracy. Thus, Examiner believes that this low probability is the result of *ignoring* very close male and female differences (i.e. classifying as independent and discarding), and has incorporated Spyros to expand upon Fig. 2 of Neti and allow for the creation of a gender-independent model (in addition to the dependent model of Neti). This would allow for further processing after gender independence is found, and to handle shorter speech segments when a gender comparison is ambiguous, or rather *creating a gender-independent phoneme model when the difference between the compared female phoneme model and the corresponding male phoneme model is less than predetermined value*. Spyro renders obvious the use of a gender-independent model when there is ambiguity in female and male models, such as in the instance when dist(i) of Neti (above equation) produces the lowest distance when a female model is compared to a male utterance in a *short* duration utterance. Without a gender-independent model created, all short

utterances having identical distances for male/female with respect to a female/male model will never be processed or will be process incorrectly as admitted by Neti above "*This method requires sufficient coverage of the phone set in the test utterance and yields poorer gender ID for short utterances. As an example, SDGid is about 62% accurate for 3-8 second utterances*".

Therefore a need exists to include a system with both dependent and independent models created as in Spyros, in order to properly handle any length utterance.

Specifically Spyros teaches a gender dependent model initially used to obtain phoneme and state alignment, wherein phonemes from several classes are used as a basis for creating the model. Creation of a *gender and context independent* phoneme model is then accomplished. Spyros also incorporates the use of a gender dependent model. Spyros then teaches that by creating and combining both gender independent and gender dependent models it prevents false alarms and unreliable gender changes, while enhancing silence detection. By utilizing both models and their benefits, segments between 3-8 seconds can be handled (Spyros 3.4).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate creating a gender-independent phoneme model when the difference between the compared female phoneme model and the corresponding male phoneme model is less than predetermined value as taught by Spyros to allow for an improved distance calculation of Neti, such as in the instance when dist(i) produces the lowest distance when a female model is compared to a male utterance in a *short* duration utterance, which results in poor gender identification or ignored comparisons, wherein Spyros improves well known differences in female and male speech to prevent false alarms and

unreliable gender changes, while enhancing silence detection (Spyros 3.4), thereby producing a steep increase in identification accuracy for short durations, whereby a *true* gender independence can be realized to successfully create a gender independent model when dependent modeling fails (e.g. distances are minimal for male and female models when comparing speech).

Further, it is obvious to one of ordinary skill in the art to apply the known technique of Spyros' dual dependent/independent gender models to the known distance based gender-dependent identification system of Neti ready for improvement to yield predictable results. Neti clearly identifies a need for improvement, i.e. increased accuracy on shorter segments, and therefore Spyros would allow for a predictable result that can handle short durations, which is when the most errors occur and when difference in gender is more negligible and harder to distinguish.

Argument (pages 13-14 section IV. and V. & VI. referring back to IV.):

- "Sukkar nowhere describes "computing accumulated confidence values for each of the plurality of data classes that the current feature vector belongs to each one of the plurality of data classes, the confidence value for each data class of the plurality of data classes based on the current best estimate for the data class and on previous confidence values for the data class, the previous confidence values associated with previous feature vectors of the audio stream," and "weighing the class-dependent phoneme models based on the accumulated confidence values," and

"recognizing the current feature vector based on the weighted class-dependent phoneme models," as set forth in claim 17. (emphases added) The Office Action makes no attempt to point out with any clarity where each of these limitations can be found in Sukkar"

Response to argument:

Examiner disagrees and maintains the use of Sukkar. Initially Examiner would like to point out that both Sukkar Neti and Sukkar are analogous with one another, such as through classification and prediction, for example the use of HMM in identification or classification.

Neti already discloses the use of phone models with gender dependency, wherein Neti teaches a method of constructing a gender-dependent speech recognition model includes the steps of providing training data of a known gender, aligning the training data, tagging the training data with a gender to create gender-tagged data, determining a gender question at a node to determine gender dependence of the gender-tagged data, determining a phonetic context question at the node to determine phonetic context dependence of the gender-tagged data, determining a highest value of an evaluation function between the gender dependence and the phonetic context dependence to determine which dependence is a dominant dependence, splitting the data of the dominant dependence into child nodes according to likelihood criteria, comparing the highest value with a threshold value to determine if additional splitting is necessary, repeating theses steps for each child node until the highest value is below the threshold value and counting the nodes having gender dependence to determine an overall

gender dependence level. A gender-dependent speech recognition system includes an input device for inputting speech to a preprocessor. The preprocessor converts the speech into acoustic data, and a processor for identifies gender-dependent phone state models and phone state modes common to both genders. The phone state models are stored in a memory device wherein the processor recognizes the speech in accordance with the phone state models (Neti Abstract & Fig. 2 and 4 below).

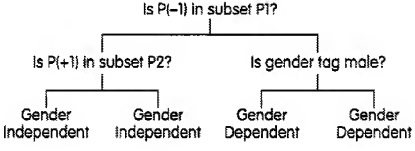

Fig. 2

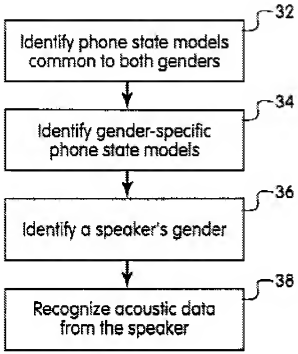

Fig. 4

Sukkar has been incorporated to teach well known used of HMM, wherein Sukkar teaches a classification system and method that post-processes HMM scores with additional confidence scores to derive a value that may be applied to a threshold on which a keyword verses non-keyword determination may be based. The first stage comprises Generalized Probabilistic Descent (GPD) analysis which uses feature vectors of the spoken words and the HMM segmentation information (developed by the HMM detector during processing) as inputs to develop a first set of confidence scores through a linear combination (a weighted sum) of the feature vectors of the speech. The second stage comprises a linear discrimination method that combines the HMM scores and the confidence scores from the GPD stage with a weighted sum to derive a second

confidence score. The output of the second stage may then be compared to a predetermined threshold to determine whether the spoken word or words includes a keyword (Sukkar Abstract).

It is well known in the art that HMM is utilized when multiple probabilistic states need to be considered, such as in classification, wherein a future state will be predicted based on past states. Therefore, Sukkar clearly renders obvious:

"*a second computing module configured to compute an accumulated confidence values for each of the plurality of data class that the current feature vector belongs to each one of the plurality of data classes, the confidence value for each data class of the plurality of data classes based on the current best estimate for the data class and on previous confidence values for the data class, the previous confidence values associated with previous feature vectors of the audio stream;*

a weighing module configured to weigh the class-dependent phoneme models based on the accumulated confidence values; and

a recognizing module configured to recognize the current feature vector (based on the weighted class-dependent phoneme models"

As previously cited, Sukkar reads upon the claim limitations, e.g. processing HMM scores with additional confidence scores to derive a value that may be applied to a threshold on which a keyword verses non-keyword determination may be based…

combining the HMM scores and the confidence scores from the GPD stage with a weighted sum to derive a second confidence score…

feature vectors of the spoken words and the HMM segmentation information (developed by the HMM detector during processing) as inputs to develop a first set of confidence scores through a linear combination (a weighted sum) of the feature vectors of the speech…

The analysis of feature vectors, HMM, and gender dependent models of Neti are improved with Sukkar's enhanced HMM feature vector, weighting, speech recognition, confidence summation, and overall classification. Sukkar renders obvious the well known *second computing module, weighing module, and recognizing module* as claimed by Neti.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate a second computing module configured to compute an accumulated confidence values for each of the plurality of data class that the current feature vector belongs to each one of the plurality of data classes, the confidence value for each data class of the plurality of data classes based on the current best estimate for the data class and on previous confidence values for the data class, the previous confidence values associated with previous feature vectors of the audio stream;

a weighing module configured to weigh the class-dependent phoneme models based on the accumulated confidence values; and

a recognizing module configured to recognize the current feature vector (based on the weighted class-dependent phoneme models as taught by Sukkar to allow for enhanced HMM feature vector, weighting, speech recognition, confidence summation,

and overall classification (Sukkar Abstract), wherein previous states will be considered in the summation of feature vectors to permit prediction of future state classification, such that the Gaussian based distance calculation for classifying speech of Neti can be further enhanced with summing of confidence scores and weighting as is well known in speech classification, for instance weighting speech vectors to emphasis important characteristics and minimizing less important characteristics with confidence as a probabilistic assurance that the data weighted is correct.

It would have also been obvious to one of ordinary skill in the art to use a known technique of classification using HMM feature vectors, weighting, speech recognition, and confidence summation to improve similar devices in the same way, such as Neti's similar speech classification system.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subj7ect matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1-16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Neti et al. US 5953701 A (hereinafter Neti) in view of "The 1998 BBN BYBLOS Primary

System applied to English and Spanish Broadcast News Transcription" Spyros et al., 1999, DARPA Broadcast News Workshop (hereinafter Spyros) and further in view of Kanevsky et al. US 6529902 (hereinafter Kanevsky).

Re claims 1, 6, and 11, Neti teaches a method for generating a speech recognition model, the method comprising:

receiving female speech training data (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, male training data, female training data, gender specific phone state models);

generating female phoneme models based on the female speech training data (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, male training data, female training data, gender specific phone state models);

receiving male speech training data (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, male training data, female training data, gender specific phone state models);

generating male phoneme models based on the male speech training data (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, male training data, female training data, gender specific phone state models);

determining a difference between each female phoneme model and each corresponding male phoneme model (Abstract, Col. 3 lines 37-49, Col. 4 lines 10-29, aligning data with gender independent data, male training data, female training data, gender specific phone state models)

However, Neti fails to teach phoneme based models and

creating a gender-independent phoneme model when the difference between the compared female phoneme model and the corresponding male phoneme model is less than predetermined value

Spyros teaches a phoneme based gender independent model creation wherein, input is segmented into two band-specific episodes using a dual-band gender-independent phoneme decoder. Each channel episode is then segmented and gender classified in one step with a dual-gender context-dependent word decoder. Further Spyros teaches Speaker-Independent (SI), Gender-Dependent (GD) bandspecific models are estimated from the training data (Spyros Section 2).

Further, Spyros teaches segments that contain mixed bands and/or genders. The solution is to detect channel changes first, and then detect gender changes within each channel, independently. In the 1998 system we used a dual-band gender independent, context independent phone-class model for band detection, trained using the following procedure... Spyros teaches training a gender independent (GI), context independent biphone class model, using the above labels (Spyros section 3.4).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate creating a gender-independent phoneme model when the difference between the compared female phoneme model and the corresponding male phoneme model is less than predetermined value as taught by Spyros to allow for an improved distance calculation of Neti, such as in the instance when dist(i) produces the lowest distance when a female model is compared to a male utterance in a *short* duration utterance, which results in poor gender identification or ignored comparisons, wherein Spyros

improves well known differences in female and male speech to prevent false alarms and *unreliable gender changes*, while enhancing silence detection (Spyros 3.4), thereby producing a steep increase in identification accuracy for short durations, whereby a *true* gender independence can be realized to successfully create a gender independent model when dependent modeling fails (e.g. distances are minimal for male and female models when comparing speech).

Further, it is obvious to one of ordinary skill in the art to apply the known technique of Spyros' dual dependent/independent gender models to the known distance based gender-dependent identification system of Neti ready for improvement to yield predictable results. Neti clearly identifies a need for improvement, i.e. increased accuracy on shorter segments, and therefore Spyros would allow for a predictable result that can handle short durations, which is when the most errors occur and when difference in gender is more negligible and harder to distinguish.

However, Neti in view of Spyros fails to teach

adding, based on at least one criteria,

one of the gender-independent phoneme model, OR

both the female phoneme model and the corresponding male phoneme model to the speech recognition model

Kanevsky can easily substitute male and female for topics when executing a Kullback-Liebler distance method, wherein Kanevsky teaches referring to FIG. 5, which

illustrates on one-way direction process of separating features belonging to different topics and topic identification via a Kullback-Liebler distance method, texts that are labeled with different topics are denoted as 501 (e.g., topic 1), 502 (e.g., topic 2), 503 (e.g., topic 3), 504 (e.g., topic N) etc. Textual features can be represented as frequencies of words, a combination of two words, a combination of three words etc. On these features, one can define metrics that allow computation of a distance between different features. For example, if topics T.sub.i give rise to probabilities P(w.sub.t.vertline.T.sub.t), where w.sub.t run all words in some vocabulary, then a distance between two topics T.sub.i and T.sub.j can be computed as #EQU13##. Using Kullback-Liebler distances is consistent with likelihood ratio criteria that are considered above, for example, in Equation (6). Similar metrics could be introduced on tokens that include T-gram words or combination of tokens, as described above. Other features reflecting topics (e.g., key words) can also be used. For every subset of k features, one can define a k dimensional vector. Then, for two different k sets, one can define a Kullback-Liebler distance using frequencies of these k sets. Using Kullback-Liebler distance, one can check which pairs of topics are sufficiently separated from each other. Topics that are close in this metric could be combined together. For example, one can find that topics related to "LOAN" and "BANKS" are close in this metric, and therefore should be combined under a new label (e.g. "FINANCE"). Also, using these metrics, one can identify in each topic domain textual feature vectors ("balls") that are sufficiently separated from other "balls" in topic domains. These "balls" are shown in FIG. 5 as

505, 506, 504, etc. When such "balls" are identified, likelihood ratios as in FIG. 1, are computed for tokens from these "balls". (Kanevsky Col. 12 lines 15-56)

Further, Kanevsky teaches another instance of detecting whether a threshold is breached and topic similarity based on training data (Kanevsky Col. 13 lines 7-12 & lines 42-45).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Spyros to incorporate adding, based on at least one criteria, one of the gender-independent phoneme model, or both the female phoneme model and the corresponding male phoneme model to the speech recognition model as taught by Kanevsky to allow for the generation of combined data models with similar context such as male and female together (e.g. LOAN and BANKS) and also isolated data such as explicit male and female data (e.g. medical and legal), wherein topics are labeled as a group of phonemes or unigrams utilizing a Kullback-Liebler distance, where one can check which pairs of topics are sufficiently separated from each other provided a subset of k features, that one can define a k dimensional vector allowing computation of a distance between different features in the form of a trained group of model (Kanevsky Col. 12 lines 15-56).

Re claims 2, 7, and 12, Neti fails to teach the method at least one computer readable medium of claim 1, wherein the at least one criteria comprises a threshold

value or an upper limit for the total number of phoneme models in the speech recognition model.

Spyros teaches the creation of three models: a triphone within-word Phoneme-Tied Mixture (PTM) model with 50 phonetic codebooks, 256 Gaussians per phone, and approximately 29K mixture weight vectors associated with the codebooks; a quinphone within-word State-Clustered-Tied Mixture (SCTM) model with approximately3500 states, 64 Gaussians per state, and 30K mixture weight vectors; and a quinphone between-word SCTM model with a similar number of parameters (Spyros section 2).

Additionally, Spyros teaches a phoneme based gender independent model creation wherein, input is segmented into two band-specific episodes using a dual-band gender-independent phoneme decoder. Each channel episode is then segmented and gender classified in one step with a dual-gender context-dependent word decoder. Further Spyros teaches Speaker-Independent (SI), Gender-Dependent (GD) bandspecific models are estimated from the training data (Spyros Section 2).

Further, Spyros teaches segments that contain mixed bands and/or genders. The solution is to detect channel changes first, and then detect gender changes within each channel, independently. In the 1998 system we used a dual-band gender independent, context independent phone-class model for band detection, trained using the following procedure... Spyros teaches training a gender independent (GI), context independent biphone class model, using the above labels (Spyros section 3.4).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate wherein the at least one

criteria comprises a threshold value or an upper limit for the total number of phoneme models in the speech recognition model as taught by Spyros to allow for multiple probabilistic models used to differentiate parameters as well as incorporating the use of gender dependent and independent phoneme decoder to create a model based on training data (Spyros Section 2), wherein the need that exists to model gender differences that are not sufficiently modeled by context-dependent variations of Neti can be improved to be phoneme based, such as through the use of Spyros in order to handle non-existent periods of silence by detecting a difference in a channel followed by a difference in gender by first applying a gender dependent model followed by a gender independent model when phonemes can not be aligned and seperated into male and female (Spyros section 3.4).

Re claims 3, 8, and 13, Neti fails to teach the method of claim 1, wherein determining the difference includes calculating a Kullback Leibler distance between the each female phoneme model and the each corresponding male phoneme model.

However, Neti fails to teach phoneme based models and

Spyros teaches a phoneme based gender independent model creation wherein, input is segmented into two band-specific episodes using a dual-band gender-independent phoneme decoder. Each channel episode is then segmented and gender classified in one step with a dual-gender context-dependent word decoder. Further

Spyros teaches Speaker-Independent (SI), Gender-Dependent (GD) bandspecific models are estimated from the training data (Spyros Section 2).

Further, Spyros teaches segments that contain mixed bands and/or genders. The solution is to detect channel changes first, and then detect gender changes within each channel, independently. In the 1998 system we used a dual-band gender independent, context independent phone-class model for band detection, trained using the following procedure… Spyros teaches training a gender independent (GI), context independent biphone class model, using the above labels (Spyros section 3.4).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate phoneme based models as taught by Spyros to allow for incorporating the use of gender dependent and independent phoneme decoder to create a model based on training data (Spyros Section 2), wherein the need that exists to model gender differences that are not sufficiently modeled by context-dependent variations of Neti can be improved to be phoneme based, such as through the use of Spyros in order to handle non-existent periods of silence by detecting a difference in a channel followed by a difference in gender by first applying a gender dependent model followed by a gender independent model when phonemes can not be aligned and seperated into male and female (Spyros section 3.4).

However, Neti in view of Spyros fails to teach determining the difference includes calculating a Kullback Leibler distance

Kanevsky teaches referring to FIG. 5, which illustrates on one-way direction process of separating features belonging to different topics and topic identification via a Kullback-Liebler distance method, texts that are labeled with different topics are denoted as 501 (e.g., topic 1), 502 (e.g., topic 2), 503 (e.g., topic 3), 504 (e.g., topic N) etc. Textual features can be represented as frequencies of words, a combination of two words, a combination of three words etc. On these features, one can define metrics that allow computation of a distance between different features. For example, if topics T.sub.i give rise to probabilities P(w.sub.t.vertline.T.sub.t), where w.sub.t run all words in some vocabulary, then a distance between two topics T.sub.i and T.sub.j can be computed as #EQU13##. Using Kullback-Liebler distances is consistent with likelihood ratio criteria that are considered above, for example, in Equation (6). Similar metrics could be introduced on tokens that include T-gram words or combination of tokens, as described above. Other features reflecting topics (e.g., key words) can also be used. For every subset of k features, one can define a k dimensional vector. Then, for two different k sets, one can define a Kullback-Liebler distance using frequencies of these k sets. Using Kullback-Liebler distance, one can check which pairs of topics are sufficiently separated from each other. Topics that are close in this metric could be combined together. For example, one can find that topics related to "LOAN" and "BANKS" are close in this metric, and therefore should be combined under a new label (e.g. "FINANCE"). Also, using these metrics, one can identify in each topic domain textual feature vectors ("balls") that are sufficiently separated from other "balls" in topic domains. These "balls" are shown in FIG. 5 as 505, 506, 504, etc. When such "balls"

are identified, likelihood ratios as in FIG. 1, are computed for tokens from these "balls". (Kanevsky Col. 12 lines 15-56)

Further, Kanevsky teaches another instance of detecting whether a threshold is breached and topic similarity based on training data (Kanevsky Col. 13 lines 7-12 & lines 42-45).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Spyros to incorporate determining the difference includes calculating a Kullback Leibler distance as taught by Kanevsky to allow for the generation of combined data models with similar context such as male and female together (e.g. LOAN and BANKS) and also isolated data such as explicit male and female data (e.g. medical and legal), wherein topics are labeled as a group of phonemes or unigrams utilizing a Kullback-Liebler distance, where one can check which pairs of topics are sufficiently separated from each other provided a subset of k features, that one can define a k dimensional vector allowing computation of a distance between different features in the form of a trained group of model (Kanevsky Col. 12 lines 15-56).

Re claims 4, 9, and 14, Neti in view of Spyros fails to teach the method of claim 3, wherein the difference is a threshold Kullback Leibler distance quantity.

Kanevsky teaches the Kullback-Leibler distance (Kanevsky Col. 5, lines 9-11) between any two topics is at least h, where h ~s some sufficiently large threshold, also

Kanevsky teaches (Kanevsky Col. 12, lines 44-47) that while using the Kullback-Leibler distance, one can check which pairs of topics are sufficiently separated from each other, and that topics that are close in this metric could be combined together).

Kanevsky also explicitly teaches how a difference is sufficient, such as classifying data groups when compared, and also creating independence from classificaiton if there is no topic discovered (Kanevsky Col. 5 lines 8-25).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti in view of Spyros to incorporate whether the model information is insignificant is based on a threshold Kullback Leibler distance quantity as taught by Kanevsky to allow for an improved language modeling for automatic speech decoding and differentiation between data groups, wherein a sufficiently large threshold indicates either separate or combinational probabilities (Kanevsky Col. 2, lines 50-52).

Re claims 5, 10, and 15, Neti teaches method of claim 1, wherein the female phoneme models, male phoneme models, and gender-independent phoneme models are Gaussian mixture models (Neti Col. 3 lines 50-67).

However, Neti fails to teach gender-independent phoneme models

Spyros teaches a phoneme based gender independent model creation wherein, input is segmented into two band-specific episodes using a dual-band gender-independent phoneme decoder. Each channel episode is then segmented and gender classified in one step with a dual-gender context-dependent word decoder. Further

Spyros teaches Speaker-Independent (SI), Gender-Dependent (GD) bandspecific models are estimated from the training data (Spyros Section 2).

Further, Spyros teaches segments that contain mixed bands and/or genders. The solution is to detect channel changes first, and then detect gender changes within each channel, independently. In the 1998 system we used a dual-band gender independent, context independent phone-class model for band detection, trained using the following procedure... Spyros teaches training a gender independent (GI), context independent biphone class model, using the above labels (Spyros section 3.4).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate gender-independent phoneme models as taught by Spyros to allow for incorporating the use of gender dependent and independent phoneme decoder to create a model based on training data (Spyros Section 2), wherein the need that exists to model gender differences that are not sufficiently modeled by context-dependent variations of Neti can be improved to be phoneme based, such as through the use of Spyros in order to handle non-existent periods of silence by detecting a difference in a channel followed by a difference in gender by first applying a gender dependent model followed by a gender independent model when phonemes can not be aligned and separated into male and female (Spyros section 3.4).

4. Claims 17-27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Neti et al. US 5953701 A (hereinafter Neti) in view of Sukkar US 5440662 A (hereinafter Sukkar).

Re claims 17, 21, and 24, Neti teaches a system for recognizing speech data from an audio stream originating from one of a plurality of data classes ([0094]), each data class having a class-dependent phoneme model, the system comprising:

a computer processor (Col. 6 lines 24-49);

a receiving module configured to receive a current feature vector of the audio stream (Col. 6 lines 24-49);

a first computing module configured to compute a current best estimates (Col. 3 lines 50-67) that the current feature vector belongs to one of the plurality of data classes (Col. 5 lines 9-21);

However, Neti fails to teach a second computing module configured to compute an accumulated confidence values for each of the plurality of data class that the current feature vector belongs to each one of the plurality of data classes, the confidence value for each data class of the plurality of data classes based on the current best estimate for the data class and on previous confidence values for the data class, the previous confidence values associated with previous feature vectors of the audio stream;

a weighing module configured to weigh the class-dependent phoneme models based on the accumulated confidence values; and

a recognizing module configured to recognize the current feature vector (based on the weighted class-dependent phoneme models; and

Sukkar teaches a two-pass classification system and method that post-processes HMM scores with additional confidence scores to derive a value that may be applied to a threshold on which a keyword verses non-keyword determination may be based. The first stage comprises Generalized Probabilistic Descent (GPD) analysis which uses feature vectors of the spoken words and the HMM segmentation information (developed by the HMM detector during processing) as inputs to develop a first set of confidence scores through a linear combination (a weighted sum) of the feature vectors of the speech. The second stage comprises a linear discrimination method that combines the HMM scores and the confidence scores from the GPD stage with a weighted sum to derive a second confidence score. The output of the second stage may then be compared to a predetermined threshold to determine whether the spoken word or words include a keyword (Sukkar Abstract).

Further, Sukkar teaches comparisons amongst all scores to produce the highest likelihood score to classify a keyword from the words present (Sukkar Col 5 lines 14-27).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate a second computing module configured to compute an accumulated confidence values for each of the plurality of data class that the current feature vector belongs to each one of the plurality of data classes, the confidence value for each data class of the plurality of data classes

based on the current best estimate for the data class and on previous confidence values for the data class, the previous confidence values associated with previous feature vectors of the audio stream;

a weighing module configured to weigh the class-dependent phoneme models based on the accumulated confidence values; and

a recognizing module configured to recognize the current feature vector (based on the weighted class-dependent phoneme models as taught by Sukkar to allow for enhanced HMM feature vector, weighting, speech recognition, confidence summation, and overall classification (Sukkar Abstract), wherein previous states will be considered in the summation of feature vectors to permit prediction of future state classification, such that the Gaussian based distance calculation for classifying speech of Neti can be further enhanced with summing of confidence scores and weighting as is well known in speech classification, for instance weighting speech vectors to emphasis important characteristics and minimizing less important characteristics with confidence as a probabilistic assurance that the data weighted is correct.

It would have also been obvious to one of ordinary skill in the art to use a known technique of classification using HMM feature vectors, weighting, speech recognition, and confidence summation to improve similar devices in the same way, such as Neti's already existent speech classification system.

Re claims 18, 22, and 25, Neti teaches the method of claim 17, wherein computing the current vector probability includes estimating a posteriori class probability for the current feature vector (Col. 2 lines 1-8))

Re claims 19, 23, and 26, Neti fails to teach the method of claim 17, wherein computing the accumulated confidence level further comprising weighing the current vector probability more than the previous vector probabilities.

Sukkar teaches a two-pass classification system and method that post-processes HMM scores with additional confidence scores to derive a value that may be applied to a threshold on which a keyword verses non-keyword determination may be based. The first stage comprises Generalized Probabilistic Descent (GPD) analysis which uses feature vectors of the spoken words and the HMM segmentation information (developed by the HMM detector during processing) as inputs to develop a first set of confidence scores through a linear combination (a weighted sum) of the feature vectors of the speech. The second stage comprises a linear discrimination method that combines the HMM scores and the confidence scores from the GPD stage with a weighted sum to derive a second confidence score. The output of the second stage may then be compared to a predetermined threshold to determine whether the spoken word or words include a keyword (Sukkar Abstract).

Further, Sukkar teaches comparisons amongst all scores to produce the highest likelihood score to classify a keyword from the words present (Sukkar Col 5 lines 14-27).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Neti to incorporate computing the accumulated confidence level further comprising weighing the current vector probability more than the previous vector probabilities as taught by Sukkar to allow for enhanced HMM feature vector, weighting, speech recognition, confidence summation, and overall classification (Sukkar Abstract), wherein previous states will be considered in the summation of feature vectors to permit prediction of future state classification, such that the Gaussian based distance calculation for classifying speech of Neti can be further enhanced with summing of confidence scores and weighting as is well known in speech classification, for instance weighting speech vectors to emphasis important characteristics and minimizing less important characteristics with confidence as a probabilistic assurance that the data weighted is correct.

It would have also been obvious to one of ordinary skill in the art to use a known technique of classification using HMM feature vectors, weighting, speech recognition, and confidence summation to improve similar devices in the same way, such as Neti's already existent speech classification system.

Re claims 20 and 27, Neti teaches the method of claim 17, further comprising determining if another feature vector is available for analysis (Col. 6 lines 24-49).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MICHAEL COLUCCI whose telephone number is (571)270-1847. The examiner can normally be reached on 9 am - 6:00 pm , Monday - Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Michael C Colucci/
Examiner, Art Unit 2626
Patent Examiner

AU 2626
(571)-270-1847
Examiner FAX: (571)-270-2847
Michael.Colucci@uspto.gov